

WHAT IS CLAIMED IS:

1. A method for formulating a set of rules representing a situation, comprising:

finding within a collection of data related to said situation a representative collection of data comprising attribute patterns and associated

5 conclusions;

forming said set of rules by:

a) comparing a selected attribute pattern to all other attribute patterns associated with conclusions different than that of said selected attribute pattern in said representative collection to
10 match irrelevant attribute elements between said selected attribute pattern and said compared attribute patterns;

b) removing said irrelevant attribute elements from said selected attribute pattern; and

repeating a) and b) for each attribute pattern in said representative
15 collection.

2. A method for formulating a set of rules according to claim 1, further comprising removing redundant rules from said set of rules.

3. A method for formulating a set of rules according to claim 1, wherein said collection of data can be chosen to increase the relative occurrence of an infrequently occurring association between a subset of said attribute patterns and said associated conclusions.

4. A method for formulating a set of rules according to claim 1, wherein:

finding said representative collection includes:

forming said representative collection with an initial attribute pattern

5 and an associated conclusion indication drawn from said collection of data;

selecting another attribute pattern from said collection of data;

a) selecting another attribute pattern from said collection of data;

b) comparing said selected attribute pattern with all attribute patterns in said representative collection;

10 c) adding said selected attribute pattern and an associated conclusion indication to said representative collection if said selected attribute pattern matches none of said attribute patterns in said representative collection;

d) adding a conclusion indication associated with said selected attribute pattern to an associated conclusion indication of a matching attribute pattern in said representative collection; and

15

repeating a) through d) until all attribute patterns in said collection of data are exhausted.

5. A method for formulating a set of rules according to claim 4, further comprising choosing a representative conclusion for each of said attribute patterns in said representative collection by identifying a predominant conclusion based on said associated conclusion indication.

6. A method for formulating a set of rules according to claim 4, further comprising selecting a representative conclusion for at least one of said attribute patterns in said representative collection based on relevant knowledge about said collection of data.

7. A method for formulating a set of rules according to claim 4, wherein said associated conclusion indication contains associated conclusions.

8. A method for formulating a set of rules according to claim 4, wherein said associated conclusion indication contains associated conclusion counts.

9. A method for formulating a set of rules according to claim 2, wherein each rule in said set of rules is expanded into a canonical form before removing said redundant rules.

10. A system for formulating a set of rules representing a situation, comprising:

a storage media containing a set of data records related to said situation;
each of said data records includes an attribute pattern and an associated
5 conclusion;

a processor operable to manipulate said set of data records to form a representative collection of attribute patterns and associated conclusions storable on said storage media;

said processor being further operable to manipulate said representative
10 collection to remove attribute elements from each of said attribute patterns that are irrelevant to said associated conclusions to form a set of rules storable on said storage media; and

said processor is further operable to remove redundant ones of said rules from said set of rules to provide a complete and consistent rule set.

11. A system for formulating a set of rules according to claim 10, further comprising:

a sample space including said set of data records;

said processor being operable to select said set of data records from
5 said sample space to increase a relative occurrence frequency of an
infrequently occurring situation.

12. A system for formulating a set of rules according to claim 10,
wherein:

said storage media contains at least one attribute pattern associated with
a plurality of conclusions; and

5 said processor is operable to select a single conclusion as a
representative conclusion from said plurality based on a specified criteria.

13. A system for formulating a set of rules according to claim 12,
wherein said specified criteria is provided by an expert.

14. A system for formulating a set of rules according to claim 10,
wherein said processor is operable to expand said set of rules into a canonical
form before said redundant ones of said rules are removed.

15. A system for formulating a set of rules according to claim 10,
wherein:

said manipulation of said representative collection includes:

a comparator module coupled to said processor and operable to provide
5 a comparison between a selected attribute pattern and all other attribute
patterns having conclusions different than that of said selected attribute
pattern; and

said processor is further operable to identify said irrelevant attribute elements in said selected attribute pattern as selected attribute elements that
10 match attribute elements in said all other attribute patterns.

16. A computer readable memory storing a program code executable to form a set of rules representing a situation, said program code comprising:

a first code section executable to find within a collection of data related to said situation a representative collection of data comprising attribute
5 patterns and associated conclusions;

a second code section executable to compare a selected attribute pattern to all other attribute patterns associated with conclusions different than that of said selected attribute pattern in said representative collection to match irrelevant attribute elements between said selected attribute pattern and said
10 compared attribute patterns;

a third code section executable to remove said irrelevant attribute elements from said selected attribute pattern; and

a fourth code section containing logic executable to repeat said second and third code sections for each attribute pattern to form a set of rules.

17. A program code according to claim 16, further comprising a fifth code section executable to remove redundant rules from said set of rules.

18. A network of interconnected computers storing program code and data for forming a set of rules representing a situation, comprising:

a storage media coupled to said network and containing a set of data records related to said situation;

5 each of said data records includes an attribute pattern and an associated conclusion;

a processor coupled to said network and operable to manipulate said set of data records to form a representative collection of attribute patterns and associated conclusions storable on said storage media;

10 said processor being further operable to manipulate said representative collection to remove attribute elements from each of said attribute patterns that are irrelevant to said associated conclusions to form a set of rules storable on said storage media; and

15 said processor is further operable to remove redundant ones of said rules from said set of rules to provide a complete and consistent rule set.

19. A method for forming a set of rules representing a situation, comprising:

finding all non-redundant fact patterns related to said situation in a data set;

5 identifying at least one attribute in each fact pattern that contributes to a respective conclusion associated with said fact pattern; and

forming said set of rules using said identified attributes and said respective associated conclusions.

20. A method according to claim 19, further comprising removing redundancies within said set of rules.

21. A method for forming a set of rules according to claim 19, wherein said data set consists of a set of records being selected to have a first conclusion in a reduced ratio with respect to a second conclusion.

22. A method for forming a set of rules according to claim 19, wherein:

each said fact pattern is associated with a group of conclusions; and
said method further comprises selecting a single conclusion from each
5 of said groups as said respective associated conclusion.

23. A method for forming a set of rules according to claim 20, wherein
said rules are expanded into a canonical form prior to removing redundancies.

24. A carrier medium containing a program code executable to form
a set of rules representing a situation, said program code comprising:

a first code section executable to find within a collection of data related
to said situation a representative collection of data comprising attribute
5 patterns and associated conclusions;

a second code section executable to compare a selected attribute pattern
to all other attribute patterns associated with conclusions different than that of
said selected attribute pattern in said representative collection to match
irrelevant attribute elements between said selected attribute pattern and said
10 compared attribute patterns;

a third code section executable to remove said irrelevant attribute
elements from said selected attribute pattern; and

a fourth code section containing logic executable to repeat said second
and third code sections for each attribute pattern to form a set of rules.
15

25. A processor operable to execute a program code from a storage
memory, said program code comprising:

a first code section executable to find, within a collection of data related
to a situation, a representative collection of data comprising attribute patterns
5 and associated conclusions;

a second code section executable to compare a selected attribute pattern to all other attribute patterns associated with conclusions different than that of said selected attribute pattern in said representative collection to match irrelevant attribute elements between said selected attribute pattern and said compared attribute patterns;

a third code section executable to remove said irrelevant attribute elements from said selected attribute pattern;

a fourth code section containing logic executable to repeat said second and third code sections for each attribute pattern to form a set of rules; and

a fifth code section executable to remove redundant rules from said set of rules.

26. A method for formulating a set of rules representing a situation, comprising:

obtaining a set of data records related to said situation, each data record containing a set of attributes and an associated conclusion;

forming a first set of mutually exclusive attribute patterns from said data records, each attribute pattern being associated with a respective conclusion group containing at least one conclusion;

maintaining a count of data records associated each conclusion in each respective conclusion group;

forming a second set of attribute patterns from said first set, each attribute pattern in said second set being associated with a preferred conclusion chosen from said respective associated conclusion group, said attribute patterns in said second set containing attributes relevant to said situation, said second set of attribute patterns being formed by:

a) creating in said second set a copy of a selected attribute pattern with an associated preferred conclusion from said first set;

b) comparing said copied selected attribute pattern to all other attribute patterns in said first set having associated preferred conclusions different from said associated preferred conclusion of said copied selected attribute pattern thereby identifying any attributes of said copied selected attribute pattern that are irrelevant to said situation;

c) removing said irrelevant attributes from said copied selected attribute pattern in said second set; and

repeating a), b) and c) for each attribute pattern in said first set to form said second set of attribute patterns comprising said set of rules.

27. A method for formulating a set of rules representing a situation according to claim 26, further comprising:

choosing as said set of data records a subset of data records from all available data records to increase a relative occurrence of an infrequently occurring conclusion.

28. A method for formulating a set of rules representing a situation, comprising:

obtaining a set of data records, each data record containing a set of attributes forming an attribute pattern and an associated conclusion;

forming from said set of data records a first set of mutually exclusive attribute patterns each associated with a conclusion group containing at least one conclusion, said first set of attribute patterns being formed by:

a) placing a copy of an initial attribute pattern and an initial associated conclusion from an initial data record into said first set of attribute patterns, said initial associated conclusion being placed in a conclusion group in said first set of attribute patterns,

15

- c) comparing said read attribute pattern to all attribute patterns of said first set of attribute patterns;

- 20

- 25

- 35

40 repeating c) through f) until all attribute patterns for said set of data records are exhausted;

selecting a representative conclusion from each of said conclusion groups as a preferred conclusion based on criteria including said conclusion counts;

45 forming a second set of attribute patterns, each associated with respective preferred conclusions, said attribute patterns in said second set containing attributes relative to said situation, said second set of attribute patterns being formed by:

g) placing a copy of a selected attribute pattern and said associated
50 preferred conclusion from said first set of attribute patterns into said second set of attribute patterns and comparing said copied selected attribute pattern to all other attribute patterns in said first set of attribute patterns having associated preferred conclusions different from said associated preferred conclusion of said copied selected attribute pattern thereby identifying any attributes of said
55 copied selected attribute pattern that are irrelevant to said situation;

h) removing said irrelevant attributes from said copied selected attribute pattern in said second set; and

repeating g) and h) for each attribute pattern in said first set of attribute patterns to form said second set of attribute patterns, said second set of
60 attribute patterns and associated preferred conclusions forming said set of rules.

29. A method for formulating a set of rules according to claim 26, wherein said preferred conclusion for each of said attribute patterns is chosen by identifying a predominant conclusion based on said count of data records.

30. A method for formulating a set of rules according to claim 26, wherein at least one of said preferred conclusions is chosen based on relevant knowledge.